学位報告

報告番号	※甲　　第　　号

主 論 文 の 要 旨

論文題目　Traceable Peer-to-Peer Record Exchange
(P2P 環境におけるトレーサブルなレコード交換に関する研究)

氏　名　LI Fengrong (李 峰 荣)

論 文 内 容 の 要 旨

With the development of high-speed network technologies and the spread use of powerful computers, peer-to-peer (P2P) technologies are getting more and more attention lately. They are actively used in various applications such as file exchange, user communication, and content distribution. One of their initial application areas was a file exchange service. Although existing file exchange services proved the potential effectiveness of P2P technologies by showing their flexibility and scalability, several important issues such as copyright violation and exchange of untrusted information occurred.

Unlike the traditional client-server architecture, a P2P network, which consists of a large number of autonomous computers (peers) and is not dependent on a specific server, enables a peer to publish information and share data with other peers without central server control. In such an environment, tracing how data is copied between peers and how data modifications are performed is not easy because data replications and modifications are performed independently by autonomous peers. This brings inconsistency in exchanged information and causes a lack of trustworthiness.

For example, when searching for images of beautiful scenes in a P2P file exchange service, trustworthiness may not be a significant problem, but when researchers exchange and share scientific information such as genome data with other researchers, the lack of trustworthiness would be a critical concern. If a

researcher is not sufficiently confident that the data was obtained from reliable sources, he or she will hesitate to use it for research purposes. Therefore, development of trustful information exchange in a P2P network has been one of the important research issues.

To ensure the trustworthiness of exchanged data, data provenance (or called lineage) which includes information about the processes and source data items that lead to its creation and current representation becomes very important in scientific databases for the validation of data.

Based on the background above, this research attempts to extend the notion of data provenance to information exchange in a P2P network and proposes a framework for traceable P2P record exchange.

In this framework, a peer can exchange structured records with a predefined schema with other peers. Records are exchanged among peers and peers can modify, store, and delete their records independently. An important feature of the framework is that it is based on the database technologies. The framework supports a tracing facility to query the lineage of the obtained records. A tracing query is described in Datalog and executed as a recursive query among cooperating peers in a P2P network. For helping a user to write a recursive tracing query, we provide a layered architecture in which the user can treat distributed information as a virtual integrated database. The abstraction layer provided by the architecture achieves a comprehensive framework for representing traceability requirements as database queries. A user can perform source determination, duplicate detection, update checking, by issuing tracing queries. Another feature of the framework is that it employs "pay-as-you-go" approach for tracing. We assume that tracing queries do not occur frequently so that it is not a wise idea to pay high maintenance cost only for the efficient tracing. When a tracing query is given, the exchange and modification histories of the queried records are collected dynamically from the related peers; distributed peers cooperatively answer it by integrating historical data maintained by them.

In this dissertation, Chapter 1 introduces the background and the research objectives. Chapter 2 reviews the previous work for each of the specific topics covered in this dissertation and shows the features of this research. Chapter 3 describes the overview of the traceable P2P record exchange system framework. The framework consists of the following three layers: the user layer offers a user interface for the record management system to each user; the global layer provides a global virtual view containing whole records in the P2P network including information for tracing; the local layer implements the global virtual view based on the cooperation of autonomous peers and each peer only maintains the minimum information related to itself. All the tracing queries are expressed in three virtual global views in the global layer. Chapter 4 introduces how to write tracing queries and presents some tracing query examples for two types of queries: ad-hoc queries and continual queries. Chapter 5 provides query processing strategies for both ad-hoc queries and continual queries. The query processing strategies of the semi-naive method and the magic set method are compared based on some experiments. In Chapter 6, the enhancement for traceable P2P record exchange framework with materialized views is analyzed. The result is that we can use materialized views to reduce query processing cost and to provide fault-tolerance for overcoming peer failures. Chapter 7 introduces the simulator developed for the traceable P2P record exchange framework and shows experimental verifications based on the simulator system in three different example P2P networks. In this chapter, we also examine the performance of the materialized views in the framework based on the experiments. Finally, Chapter 8 concludes this dissertation.